Erratum: Limit theorems for Betti numbers of random simplicial complexes

Matthew Kahle and Elizabeth Meckes

April 24, 2014

Abstract

We correct the proofs of the main theorems in our earlier paper "Limit theorems for Betti numbers of random simplicial complexes." We are grateful to D. Yogeshwaran for pointing out the mistakes.

1 The Erdős-Renyi random clique complex

In the paper [2], we claimed a central limit theorem for the Betti number of an Erdős-Renyi random simplicial complex (Theorem 2.4). The proof given there contains an error, however, with minor modifications and an additional recent result, the proof goes through essentially as before. For brevity, we refer to [2] for notation and descriptions of the models used. Formally, we have the following modification of Theorem 2.4 of [2].

Theorem 1.1. *Consider the Erdős-Renyi clique complex $X(n,p)$; that is, take a random 1-skeleton on n vertices, in which edges are present independently and with probability p, and let $X(n,p)$ be the maximal complex over this 1-skeleton. Suppose that there is some $\delta > 0$ such that $p = \omega(n^{-1/k+\delta})$ and $p = o(n^{-1/(k+1)-\delta})$. Then*

$$\frac{\beta_k(X(n,p)) - \mathbb{E}[\beta_k(X(n,p))]}{\sqrt{\mathrm{Var}[\beta_k]}} \Rightarrow \mathcal{N}(0,1).$$

Note that the range of p is slightly restricted relative to what was claimed earlier, when δ was taken to be 0.

The mistake in the proof given in [2] of this result was the claim that, given the Morse inequalities

$$f_k - f_{k+1} - f_{k-1} \le \beta_k \le f_k,$$

and central limit theorems for the recentered, renormalized upper and lower bounds, a central limit theorem for β_k itself followed; this is not true, however, because the difference in means of the upper and lower bounds is too large relative to the normalization to allow such a conclusion. However, the proof of the central limit theorem for the lower bound is valid, and with minor modifications, one can use the same proof to obtain a central limit theorem for the quantity

$$\tilde{\beta}_k := f_k - f_{k+1} - f_{k-1} + f_{k+2} + f_{k-2} - f_{k+3} - f_{k-3} + \cdots.$$

A consequence of the results in [1] is that for p in the given regime, a.a.s. all the Betti numbers except for β_k are zero. It then follows immediately from the Euler formula that $\beta_k = \tilde{\beta}_k$ a.a.s.

A central limit theorem for β_k is then essentially immediate from a central limit theorem for $\tilde{\beta}_k$:

$$\mathbb{P}\left[\frac{\beta_k - \mathbb{E}\tilde{\beta}_k}{\sqrt{\mathrm{Var}(\tilde{\beta}_k)}} \leq t\right] \leq \mathbb{P}\left[\frac{\tilde{\beta}_k - \mathbb{E}\tilde{\beta}_k}{\sqrt{\mathrm{Var}(\tilde{\beta}_k)}} \leq t, \beta_k = \tilde{\beta}_k\right] + \mathbb{P}[\beta_k \neq \tilde{\beta}_k] \leq \mathbb{P}\left[\frac{\tilde{\beta}_k - \mathbb{E}\tilde{\beta}_k}{\sqrt{\mathrm{Var}(\tilde{\beta}_k)}} \leq t\right] + \mathbb{P}[\beta_k \neq \tilde{\beta}_k];$$

the corresponding lower bound follows the same way. Thus

$$\left|\mathbb{P}\left[\frac{\beta_k - \mathbb{E}\tilde{\beta}_k}{\sqrt{\mathrm{Var}(\tilde{\beta}_k)}} \leq t\right] - \Phi(t)\right| \leq \left|\mathbb{P}\left[\frac{\tilde{\beta}_k - \mathbb{E}\tilde{\beta}_k}{\sqrt{\mathrm{Var}(\tilde{\beta}_k)}} \leq t\right] - \Phi(t)\right| + \mathbb{P}[\beta_k \neq \tilde{\beta}_k]. \tag{1}$$

For k fixed, the second quantity tends to zero as $n \to \infty$ and if $\tilde{\beta}_k$ satisfies a central limit theorem, then so does the first quantity, and we are done.

To move to the actual statement of Theorem 1.1, one needs a slightly more refined version of the a.a.s. equality of β and $\tilde{\beta}_k$. In fact, the techniques in [1] give that for p in the given regime, $\beta_k = \tilde{\beta}_k$ with probability $1 - o\left(n^{-M}\right)$ for any constant $M > 0$. On the other hand, since a simplicial complex on n vertices has $\beta_k \leq f_k \leq \binom{n}{k+1}$, in all cases we have that $\beta_k - \tilde{\beta}_k = O\left(n^{k+1}\right)$. It is shown below that $\mathrm{Var}(\tilde{\beta}_k) \sim n^{2k}p^{2\binom{k+1}{2}-1}$. These estimates together are enough to obtain Theorem 1.1, as follows. First note that the above estimates imply that

$$\frac{|\mathbb{E}\beta_k - \mathbb{E}\tilde{\beta}_k|}{\sqrt{\mathrm{Var}(\tilde{\beta}_k)}} = o\left(n^{1-M}p^{\frac{1}{2}-\binom{k+1}{2}}\right);$$

choosing $M \geq \frac{k+3}{2}$, we have that in the regime of p specified in the theorem, $\frac{|\mathbb{E}\beta_k - \mathbb{E}\tilde{\beta}_k|}{\sqrt{\mathrm{Var}(\tilde{\beta}_k)}} \to 0$ as $n \to \infty$. Now,

$$\frac{\mathrm{Var}(\beta_k)}{\mathrm{Var}(\tilde{\beta}_k)} = 1 + \frac{\mathrm{Var}(\beta_k - \tilde{\beta}_k)}{\mathrm{Var}(\tilde{\beta}_k)} + 2\frac{\mathrm{Cov}(\tilde{\beta}_k, \beta_k - \tilde{\beta}_k)}{\mathrm{Var}(\tilde{\beta}_k)}.$$

By the Cauchy-Schwarz inequality, $\frac{|\mathrm{Cov}(\tilde{\beta}_k, \beta_k - \tilde{\beta}_k)|}{\mathrm{Var}(\tilde{\beta}_k)} \leq \sqrt{\frac{\mathrm{Var}(\beta_k - \tilde{\beta}_k)}{\mathrm{Var}(\tilde{\beta}_k)}}$, and

$$\frac{\mathrm{Var}(\beta_k - \tilde{\beta}_k)}{\mathrm{Var}(\tilde{\beta}_k)} \leq \frac{\mathbb{E}|\beta_k - \tilde{\beta}_k|^2}{\mathrm{Var}(\tilde{\beta}_k)} = o\left(n^{2-M}p^{1-2\binom{k+1}{2}}\right),$$

which tends to zero if $M \geq k + 3$. We thus have that $\frac{\mathrm{Var}(\beta_k)}{\mathrm{Var}(\tilde{\beta}_k)} \to 1$ as $n \to \infty$. Finally, we get that for any $\delta, \delta', \epsilon > 0$, for n large enough,

$$\mathbb{P}\left[\frac{\beta_k - \mathbb{E}\beta_k}{\sqrt{\mathrm{Var}(\beta_k)}} \leq t\right] = \mathbb{P}\left[\frac{\beta_k - \mathbb{E}\tilde{\beta}_k}{\sqrt{\mathrm{Var}(\tilde{\beta}_k)}} \leq t\sqrt{\frac{\mathrm{Var}(\beta_k)}{\mathrm{Var}(\tilde{\beta}_k)}} + \left(\frac{\mathbb{E}\beta_k - \mathbb{E}\tilde{\beta}_k}{\sqrt{\mathrm{Var}(\tilde{\beta}_k)}}\right)\right]$$

$$\leq \mathbb{P}\left[\frac{\beta_k - \mathbb{E}\tilde{\beta}_k}{\sqrt{\mathrm{Var}(\tilde{\beta}_k)}} \leq t(1+\delta) + \delta'\right]$$

$$\leq \Phi(t(1+\delta) + \delta') + \epsilon.$$

Letting $\delta, \delta', \epsilon$ tend to zero, we have

$$\limsup_{n\to\infty} \mathbb{P}\left[\frac{\beta_k - \mathbb{E}\beta_k}{\sqrt{\mathrm{Var}(\beta_k)}} \leq t\right] \leq \Phi(t).$$

The lower bound is proved the same way, and thus it suffices to prove a central limit theorem for $\tilde{\beta}_k$. To do this, we follow essentially the same proof as the one given in [2] for $f_k - f_{k-1} - f_{k+1}$. In particular, the result is an application of the following result of Barbour, Karoński, and Ruciński.

Theorem 1.2. *Let $\{X_{\mathbf{j}} : \mathbf{j} = (j_1, \ldots, j_r) \in J\}$ be a dissociated set of random variables, such that $\mathbb{E}X_{\mathbf{j}} = 0$ for all $\mathbf{j}$. Let $W := \sum_{\mathbf{j} \in J} X_{\mathbf{j}}$ and suppose that the $X_{\mathbf{j}}$ are normalized such that $\mathbb{E}W^2 = 1$. Then*

$$d_1(W, Z) \leq K \sum_{\mathbf{j} \in J} \sum_{\mathbf{k}, \mathbf{l} \in L_{\mathbf{j}}} \left[\mathbb{E}|X_{\mathbf{j}} X_{\mathbf{k}} X_{\mathbf{l}}| + \mathbb{E}|X_{\mathbf{j}} X_{\mathbf{k}}| \mathbb{E}|X_{\mathbf{l}}| \right], \tag{2}$$

where Z is a standard normal random variable and $L_{\mathbf{j}}$ is the dependency neighborhood of $\mathbf{j}$.

Write

$$\tilde{\beta}_k = \sum_{A \subseteq \mathcal{V}} (-1)^{|A|+k+1} \xi_A,$$

where $\mathcal{V}$ is the collection of n vertices over which our complex is built and ξ_A is the indicator that A spans a complete graph in the random complex $X(n,p)$; that is, that all $\binom{|A|}{2}$ potential edges between vertices in A are present. Let $\sigma^2 := \mathrm{Var}\left(\tilde{\beta}_k\right)$, and consider the random variable

$$W := \frac{\tilde{\beta}_k - \mathbb{E}\tilde{\beta}_k}{\sqrt{\mathrm{Var}(\tilde{\beta}_k)}} = \frac{1}{\sigma} \sum_{A \subseteq \mathcal{V}} (-1)^{|A|+k+1} (\xi_A - \mathbb{E}\xi_A);$$

that is,

$$X_A := \frac{(-1)^{|A|+k+1}}{\sigma} (\xi_A - \mathbb{E}\xi_A).$$

It is not hard to see that for any subsets A, B, C,

$$\mathbb{E}|X_A X_B X_C| + \mathbb{E}|X_A X_B| \mathbb{E}|X_C| \leq \frac{16}{\sigma^3} \mathbb{E}\left[\xi_A \xi_B \xi_C\right],$$

and it thus suffices to estimate

$$\frac{16}{\sigma^3} \sum_{A \subseteq \mathcal{V}} \sum_{B, C \in L_A} \mathbb{E}\left[\xi_A \xi_B \xi_C\right],$$

where for $A \subseteq \mathcal{V}$, L_A is the collection of subsets of $\mathcal{V}$ sharing at least two vertices with A (so that they have at least one potential edge in common). Decomposing by the sizes of A, B, C and the sizes of their intersections, we have that

$$\frac{16}{\sigma^3} \sum_{A \subseteq \mathcal{V}} \sum_{B, C \in L_A} \mathbb{E}\left[\xi_A \xi_B \xi_C\right]$$

$$= \frac{16}{\sigma^3} \sum_{\substack{\ell_A, \ell_B, \ell_C \geq 2 \\ r_{A,B} \geq 2 \\ r_{A \backslash B, C} \geq 0 \\ r_{A \cap B, C} \geq (2 - r_{A \backslash B, C})_+ \\ r_{B \backslash A, C} \geq 0}} \mathrm{C}\, p^{\binom{\ell_A}{2} + \binom{\ell_B}{2} + \binom{\ell_C}{2} - \binom{r_{A,B}}{2} - \binom{r_{A \backslash B, C} + r_{A \cap B, C}}{2} - \binom{r_{B \backslash A, C} + r_{A \cap B, C}}{2} + \binom{r_{A \cap B, C}}{2}},$$

where the upper limits all depend only on k, and the combinatorial coefficient C is given by

$$\mathrm{C} = \binom{n}{\ell_A} \binom{n - \ell_A}{\ell_B - r_{A,B}} \binom{\ell_A}{r_{A,B}} \binom{\ell_B - r_{A,B}}{r_{B \backslash A, C}}$$

$$\times \binom{r_{A,B}}{r_{A \cap B, C}} \binom{\ell_A - r_{A,B}}{r_{A \backslash B, C}} \binom{n - \ell_A - \ell_B + r_{A,B}}{\ell_C - r_{A \backslash B, C} - r_{B \backslash A, C} - r_{A \cap B, C}}$$

$$\leq c_k n^{\ell_A + \ell_B + \ell_C - r_{A,B} - r_{A \backslash B, C} - r_{B \backslash A, C} - r_{A \cap B, C}},$$

3

for a constant c_k depending only on k. If we fix $\ell_A, \ell_B, r_{A,B}$ and ignore for the moment those factors that depend only on these parameters, we are left with sums over ℓ_C, etc. of terms of size

$$\frac{1}{\sigma^3} n^{\ell_C - r_{A\setminus B,C} - r_{B\setminus A,C} - r_{A\cap B,C}} p^{\binom{\ell_C}{2} - \binom{r_{A\setminus B,C} + r_{A\cap B,C}}{2} - \binom{r_{B\setminus A,C} + r_{A\cap B,C}}{2} + \binom{r_{A\cap B,C}}{2}}. \tag{3}$$

Now, if ℓ_C is increased by one and the new element of C is also in, say $A\setminus B$, then the power of n in the expression above does not change, but the power of p does; the ratio of the new term to the old is

$$p^{\binom{\ell_C+1}{2} - \binom{\ell_C}{2} - \binom{r_{A\setminus B,C}+1+r_{A\cap B,C}}{2} + \binom{r_{A\setminus B,C}+r_{A\cap B,C}}{2}} = p^{\ell_c - r_{A\setminus B,C} - r_{A\cap B,C}}.$$

Similarly, if ℓ_C is increased by one and the new element of C is also in $A\cap B$, then the ratio of the new term to the old is

$$p^{\ell_C - r_{A\setminus B,C} - r_{B\setminus A,C} - r_{A\cap B,C}}.$$

Since in both cases the power on p is nonnegative, adding a new vertex to C which is already in $A\cup B$ can only make the summand smaller. On the other hand, if ℓ_C is increased by 1, and the new vertex is not in A or B, then the ratio of the new term to the old is np^{ℓ_C}. In the regime that we consider, this tends to infinity for (the old) $\ell_C \le k$ and tends to zero for $\ell_C \ge k+1$; that is, the largest possible order for the expression in (3) is achieved when $\ell_C = k+1$, when $r_{A\setminus B,C} + r_{A\cap B,C} = 2$ and $r_{B\setminus A,C} = 0$. Using these values in (3) yields

$$\frac{1}{\sigma^3} n^{k-1} p^{\binom{k+1}{2}-1}. \tag{4}$$

Now suppose that only ℓ_A is fixed, and ignore the part of the summand depending only on its value. We thus must consider summands of the size

$$\frac{1}{\sigma^3} n^{\ell_B - r_{A,B} + k - 1} p^{\binom{\ell_B}{2} - \binom{r_{A,B}}{2} + \binom{k+1}{2}-1}. \tag{5}$$

As before, if ℓ_B is increased and so is $r_{A,B}$, then the expression can only get smaller. If ℓ_B is increased by 1 while $r_{A,B}$ stays fixed, then the ratio of the new expression to the old is np^{ℓ_B}, and so we once again see that the largest possible size of the expression in (5) is achieved when $\ell_B = k+1$ and $r_{A,B} = 2$; the quantity in (5) is thus bounded above by

$$\frac{1}{\sigma^3} n^{2k-2} p^{2\binom{k+1}{2}-2}. \tag{6}$$

Finally, considering the full term, we have the upper bound of

$$\frac{1}{\sigma^3} n^{\ell_A + 2k - 2} p^{\binom{\ell_A}{2} + 2\binom{k+1}{2}-2}; \tag{7}$$

by the same argument one last time, this expression is maximized when $\ell_A = k+1$, yielding

$$\frac{1}{\sigma^3} n^{3k-1} p^{3\binom{k+1}{2}-2}; \tag{8}$$

that is, Theorem 2 implies that

$$d_1(W, Z) \le \frac{C}{\sigma^3} n^{3k-1} p^{3\binom{k+1}{2}-2},$$

where $W = \frac{\tilde{\beta}_k - \mathbb{E}\tilde{\beta}_k}{\sqrt{\mathrm{Var}(\tilde{\beta}_k)}}$.

The computation of σ^2 from [2] essentially goes through as before. It was shown there that

$$\mathrm{Var}(f_k) \sim c_k n^{2k} p^{2\binom{k+1}{2}-1}$$

4

(the numbered subclaim and equation (4) of [2] are inconsistent and in fact both wrong: unfortunate casualties of a change of index in the course of editing). From this it follows that for any $j > 0$,

$$\frac{\text{Var}(f_{k\pm j})}{\text{Var}(f_k)} \to 0;$$

moreover, one can compute covariances as in [2], yielding for example the formula

$$\text{Cov}(f_{k+j}, f_{k+\ell}) = \binom{n}{k+j+1} p^{\binom{k+j+1}{2}+\binom{k+\ell+1}{2}} \sum_{r=2}^{k+j+1} \binom{k+j+1}{r} \binom{n-k-j-1}{k+\ell+1-r} \left(p^{-\binom{r}{2}} - 1\right),$$

for $0 \le j \le k$ (and similarly in other cases). Again one confirms that the order of this expression is smaller than the order of the variance of f_k, so that we finally have that

$$\sigma^2 \sim n^{2k} p^{2\binom{k+1}{2}-1}.$$

The sums over indices only contribute constants depending on k, so that we have that the error in the abstract normal approximation theorem above is bounded above by

$$c_k \frac{n^{3k-1} p^{3\binom{k+1}{2}-2}}{n^{3k} p^{3\binom{k+1}{2}-\frac{3}{2}}} = \frac{c_k}{n\sqrt{p}},$$

for a constant c_k depending only on k.

2   The Čech complex

In [2] we claimed three different limit theorems for the kth Betti number of a random Čech complex: depending on the sub-regime of the sparse regime, the kth Betti number either vanished a.a.s., had an approximate Poisson distribution, or satisfied a central limit theorem. The approach taken in [2] works in most of the sparse regime, namely as long as $n^{k+3} r_n^{d(k+2)} \to 0$, but to deal with the regime in which $r_n = o(n^{-1/d-\delta})$ for some $\delta > 0$, but $n^{k+3} r_n^{d(k+2)}$ is bounded away from zero, a slightly different argument is needed, for the same reason as in the previous section.

We begin by noting that one can write β_k semi-explicitly as follows. Let S_k denote the number of empty k+1-dimensional simplex components of the Čech complex $\mathcal{C}(X_1, \dots, X_n)$ spanned by $X_1, \dots, X_n$. Note that every such connected component has exactly $k + 2$ vertices.

For every pair of integers $i > k+2$ and $j > 0$, let $X_{i,j}$ denote the number of connected components C of $\mathcal{C}(X_1, \dots, X_n)$ on i vertices such that $\beta_k(C) = j$. In other words $X_{i,j}$ counts the components on i vertices which contribute exactly j to β_k.

Then

$$\beta_k = S_k + \sum_{i>k+2, j>0} j X_{i,j}.$$

A central limit theorem for β_k is a indeed a consequence of a central limit theorem for S_k as claimed in [2], by a slightly more careful analysis.

Set $m = \lfloor 1 + 1/(\delta d) \rfloor$, and define the truncated sum

$$\tilde{\beta}_k = S_k + \sum_{i=k+3}^{m} \sum_{j>0} j X_{i,j}.$$

By a modification of the argument in [2], one obtains the following.

Theorem 2.1. *With notation as above, for $r_n = o\left(n^{-1/d-\delta}\right)$ and $\lim_{n\to\infty} n^{k+2} r_n^{d(k+1)} \to \infty$,*

$$\frac{1}{\sqrt{n(nr_n^d)^{k+1}}}\left(\tilde{\beta}_k - \mathbb{E}\left[\tilde{\beta}_k\right]\right) \Longrightarrow \mathcal{N}\left(0, \frac{\mu_{k+2,1}}{(k+2)!}\right),$$

where $\mu_{k+2,1}$ is a constant depending only on f and k.

From here, a central limit theorem for β_k itself follows:

Theorem 2.2. *With notation as above, for $r_n = o\left(n^{-1/d-\delta}\right)$ and $\lim_{n\to\infty} n^{k+2} r_n^{d(k+1)} \to \infty$,*

$$\frac{1}{\sqrt{n(nr_n^d)^{k+1}}}\left(\beta_k - \mathbb{E}\left[\beta_k\right]\right) \Longrightarrow \mathcal{N}\left(0, \frac{\mu_{k+2,1}}{(k+2)!}\right),$$

where $\mu_{k+2,1}$ is the same constant as in Theorem 2.1.

Compare with Theorem 3.2 (iii) of [2]: the range of r_n is slightly more restricted (there, δ was taken to be 0); the theorem here is also stated in terms of a specific numerical normalization, rather than abstractly in terms of the variance of β_k as in [2].

Proof of Theorem 2.2 from Theorem 2.1. Observe that

$$\mathbb{P}\left[\frac{\beta_k - \mathbb{E}\beta_k}{\sqrt{n(nr_n^d)^{k+1}}} \le t\right] \le \mathbb{P}\left[\frac{\tilde{\beta}_k - \mathbb{E}\tilde{\beta}_k}{\sqrt{n(nr_n^d)^{k+1}}} \le t+\epsilon\right] + \mathbb{P}\left[\left|\frac{\tilde{\beta}_k - \beta_k - \mathbb{E}[\tilde{\beta}_k - \beta_k]}{\sqrt{n(nr_n^d)^{k+1}}}\right| > \epsilon\right]$$

$$\le \mathbb{P}\left[\frac{\tilde{\beta}_k - \mathbb{E}\tilde{\beta}_k}{\sqrt{n(nr_n^d)^{k+1}}} \le t+\epsilon\right] + \frac{2\mathbb{E}|\beta_k - \tilde{\beta}_k|}{\epsilon\sqrt{n(nr_n^d)^{k+1}}}.$$

Claim: $\dfrac{\mathbb{E}|\beta_k - \tilde{\beta}_k|}{\sqrt{n(nr_n^d)^{k+1}}} \xrightarrow{n\to\infty} 0.$

From the claim it follows that, given $\epsilon > 0$, there is an n large enough so that

$$\mathbb{P}\left[\frac{\beta_k - \mathbb{E}\beta_k}{\sqrt{n(nr_n^d)^{k+1}}} \le t\right] \le \mathbb{P}\left[\frac{\tilde{\beta}_k - \mathbb{E}\tilde{\beta}_k}{\sqrt{n(nr_n^d)^{k+1}}} \le t+\epsilon\right] + \epsilon.$$

Using the central limit theorem already established for $\tilde{\beta}_k$ and then letting $\epsilon \to 0$ shows that

$$\limsup_{n\to\infty} \mathbb{P}\left[\frac{\beta_k - \mathbb{E}\beta_k}{\sqrt{n(nr_n^d)^{k+1}}} \le t\right] \le \mathbb{P}\left[\sqrt{\frac{\mu_{k+2,1}}{(k+2)!}} Z \le t\right],$$

where Z is a standard Gaussian random variable. The corresponding lower bound follows in the same way, so that given the claim, the proof of Theorem 2.2 is complete.

To prove the claim, observe that

$$|\beta_k - \tilde{\beta}_k| = \sum_{i=m+1}^{S}\sum_{j>0} jX_{i,j} + \sum_{i=S+1}^{n}\sum_{j>0} X_{i,j}, \tag{9}$$

where $S = \left\lceil \frac{2}{d\delta} + 1 \right\rceil$.

Since there are i^{i-2} spanning trees on a set of i vertices, and since a connected component of order i can contribute at most $\binom{i}{k+1}$ to β_k, we have that for fixed $i \ge m+1$,

$$\mathbb{E}\left[\sum_{j>0} jX_{i,j}\right] \le \binom{n}{i} i^{i-2} \left(\|f\|_\infty r_n^d\right)^{(i-1)} \binom{i}{k+1} \le \frac{n^i}{i!} i^{i-2} \left(\|f\|_\infty r_n^d\right)^{(i-1)} \binom{i}{k+1}.$$

6

It follows that

$$2S \sum_{i \geq m+1}^{S} \mathbb{E}\left[\sum_{j>0} jX_{i,j}\right] = O\left(n^{m+1}r_n^{dm}\right) = O\left(n^{-d\delta}\right),$$

since $r_n = o(n^{-1/d-\delta})$; this takes care of the first sum in (9).

For the second sum, the same estimate on the terms gives that

$$\sum_{i=S+1}^{n} \mathbb{E}\left[\sum_{j>0} jX_{i,j}\right] = O\left(n^2(nr_n^d)^S\right) = O\left(n^{2-d\delta S}\right).$$

Since $S > \frac{2}{d\delta} + 1$, we have that $\mathbb{E}\left|\beta_k - \tilde{\beta}_k\right| = O\left(n^{-\delta d}\right)$, which proves the claim.

$\square$

As in [2], to prove Theorem 2.1, we consider the Poissonized problem first, then recover the i.i.d. case.

Let N_n be a Poisson random variable with mean n, and let $\mathcal{P}_n = \{X_1, \ldots, X_{N_n}\}$, where $\{X_i\}_{i=1}^\infty$ is an i.i.d. sequence of random points in $\mathbb{R}^d$ with density f. Then $\mathcal{P}_n$ is a Poisson process with intensity $nf(\cdot)$, and one can define S_k^P and $X_{i,j}^P$ for the random points $\mathcal{P}_n$ analogously to the earlier definitions. In what follows, assume that $k \geq 2$; that is, the empty k-simplices are at least empty triangles. Empty 1-simplices are simply pairs of vertices which are not connected, and different arguments are needed in that case.

In order to compute expectations for the expressions which arise in the Poissonized case, the following results are useful.

Theorem 2.3 (See Theorem 1.6 of [3])**.** *Let $\lambda > 0$ and let $\mathcal{P}_\lambda$ be a Poisson process with intensity $\lambda f(\cdot)$. Let $j \in \mathbb{N}$, and suppose that $h(\mathcal{Y}, \mathcal{S})$ is a bounded measurable function on pairs $(\mathcal{Y}, \mathcal{S})$ with $\mathcal{S}$ a finite subset of $\mathbb{R}^d$ and $\mathcal{Y} \subseteq \mathcal{S}$, such that $h(\mathcal{Y}, \mathcal{S}) = 0$ unless $|\mathcal{Y}| = j$. Then*

$$\mathbb{E}\left[\sum_{\mathcal{Y} \subseteq \mathcal{P}_\lambda} h(\mathcal{Y}, \mathcal{P}_\lambda)\right] = \frac{\lambda^j}{j!} \mathbb{E}h(\mathcal{X}_j, \mathcal{X}_j \cup \mathcal{P}_\lambda),$$

where $\mathcal{X}_j$ is a set of j i.i.d. points in $\mathbb{R}^d$ with density f, independent of $\mathcal{P}_\lambda$.

From this, one can prove the following (see [2] for the proof).

Theorem 2.4. *Let $\lambda > 0$ and $k, j_1, \ldots, j_k \in \mathbb{N}$; define $j := \sum_{i=1}^k j_i$. For $1 \leq i \leq k$, suppose $h_i(\mathcal{Y}, \mathcal{S})$ is a bounded measurable function of pairs $(\mathcal{Y}, \mathcal{S})$ of finite subsets of $\mathbb{R}^d$ with $\mathcal{Y} \subseteq \mathcal{S}$, such that $h_i(\mathcal{Y}, \mathcal{S}) = 0$ if $|\mathcal{Y}| \neq j_i$. Then*

$$\mathbb{E}\left[\sum_{\mathcal{Y}_1 \subseteq \mathcal{P}_\lambda} \cdots \sum_{\mathcal{Y}_k \subseteq \mathcal{P}_\lambda} \left(\prod_{i=1}^k h_i(\mathcal{Y}_i)\right) \mathbb{1}_{\{\mathcal{Y}_i \cap \mathcal{Y}_j = \emptyset \text{ for } i \neq j\}}\right] = \mathbb{E}\left[\prod_{i=1}^k \left(\frac{\lambda^{j_i}}{j_i!}\right) h_i\left(\mathcal{X}_{j_i}, \left(\cup_{i=1}^k \mathcal{X}_{j_i}\right) \cup \mathcal{P}_n\right)\right],$$

where $\mathcal{X}_{j_i}$ are j_i i.i.d points in $\mathbb{R}^d$ with density f, $\mathcal{P}_\lambda$ is a Poisson process with intensity $\lambda f(\cdot)$, and $\{\mathcal{X}_{j_i}\}_{i=1}^k$ and $\mathcal{P}_\lambda$ are all independent.

One can apply these results to compute the mean and variance of $\tilde{\beta}_{k,A}^P$, the contribution to $\tilde{\beta}_k^P$ from components whose left-most vertex is in an open set A with $\text{vol}(\partial A) = 0$.

In order to apply the lemmas, the corresponding means in the i.i.d. case are needed.

7

Lemma 2.5. *Let $g_{i,j,A}(x_1, \ldots, x_i)$ be the indicator that $\mathcal{C}(\{x_1, \ldots, x_i\})$ is connected, has kth Betti number equal to j, and has left-most-point in A. Then for $\{X_i\}$ i.i.d. with density f as before, there is a constant $\mu_{i,j,A}$ depending only on i, j, f, and A such that*

$$\lim_{n \to \infty} r_n^{-d(i-1)} \mathbb{E}\left[g_{i,j,A}(X_1, \ldots, X_i)\right] = \mu_{i,j,A}.$$

The proof is identical to that of the analogous result in Chapter 3 of Penrose [3].

Lemma 2.6. *For $\mu_{i,j,A}$ as in Lemma 2.5,*

(i)
$$\lim_{n \to \infty} n^{-(k+2)} r_n^{-d(k+1)} \mathbb{E}\left[\tilde{\beta}_k^P\right] = \lim_{n \to \infty} n^{-(k+2)} r_n^{-d(k+1)} \mathrm{Var}\left[\tilde{\beta}_k^P\right] = \frac{\mu_{k+2,1,A}}{(k+2)!}.$$

(ii)
$$\lim_{n \to \infty} n^{-(k+2)} r_n^{-d(k+1)} \mathbb{E}\left[S_k^P\right] = \lim_{n \to \infty} n^{-(k+2)} r_n^{-d(k+1)} \mathrm{Var}\left[S_k^P\right] = \frac{\mu_{k+2,1,A}}{(k+2)!}.$$

(iii)

$$\lim_{n \to \infty} n^{-(k+3)} r_n^{-d(k+2)} \mathbb{E}\left[\sum_{\substack{k+3 \le i \le m \\ j \ge 0}} j X_{i,j}\right]$$

$$= \lim_{n \to \infty} n^{-(k+3)} r_n^{-d(k+2)} \mathrm{Var}\left[\sum_{\substack{k+3 \le i \le m \\ j \ge 0}} j X_{i,j}\right] = \frac{1}{(k+3)!} \sum_{j=1}^{\binom{k+3}{k+1}} j \mu_{k+3,j,A}.$$

Proof. For $i \ge k+2$ and $j \ge 1$, let $h_{i,j,A}(\{x_0, \ldots, x_k\}, \mathcal{X})$ be the indicator that $\{x_0, \ldots, x_{i-1}\} \subseteq \mathcal{X}$ form a connected component of $\mathcal{C}(\mathcal{X})$ with $\tilde{\beta}_k(\mathcal{C}(x_0, \ldots, x_{i-1})) = j$, whose left-most point is in A. Then with $\tilde{\beta}_{k,A}^P$ denoting the sub-sum of $\tilde{\beta}_k$ coming from those components with left-most point in A,

$$\mathbb{E}[\tilde{\beta}_{k,A}^P] = \mathbb{E}\left[\sum_{\substack{k+2 \le i \le m \\ j \ge 1}} \sum_{\mathcal{Y} \subseteq \mathcal{P}_\lambda} j h_{i,j,A}(\mathcal{Y}, \mathcal{P}_n)\right] = \sum_{\substack{k+2 \le i \le m \\ j \ge 1}} \frac{n^i}{i!} \mathbb{E}\left[j h_{i,j,A}(\mathcal{X}_i, \mathcal{X}_i \cup \mathcal{P}_n)\right]. \quad (10)$$

Now, $\mathbb{E}\left[h_{i,j,A}(\mathcal{X}_i, \mathcal{X}_i \cup \mathcal{P}_n)\right] \le \mathbb{E}\left[g_{i,j,A}(\mathcal{X}_i)\right]$, where $g_{i,j,A}(\mathcal{X}_i)$ is the indicator that the i i.i.d. points $\mathcal{X}_i$ are connected (with respect to cut-off radius r_n) with kth Betti number of the complex they span equal to j (ignoring any issues of connectedness to anything else). By Lemma 2.5 $\mathbb{E}\left[g_{i,j,A}(\mathcal{X}_i)\right] \simeq r_n^{d(i-1)} \mu_{i,j,A}$. Note moreover that the conditional probability that $\mathcal{X}_i$ is isolated from $\mathcal{P}_n$ given that $\mathcal{X}_i$ is connected and has left-most vertex in A is bounded below by the probability that there are no points of $\mathcal{P}_n$ in the ball of radius $2(\ell_{i,j} + 1)r_n$ about X_1, where $\ell_{i,j}$ is the largest number of edges that may be needed to move from one vertex to another in a simplicial complex on i vertices with kth Betti number equal to j. Since $\mathcal{P}_n$ is a Poisson process with intensity $nf(\cdot)$, this probability is given exactly by $e^{-n\,\mathrm{vol}_f(B_{2(\ell_{i,j}+1)r_n}(X_1))} \ge e^{-n\|f\|_\infty \theta_d(2(\ell_{i,j}+1)r_n)^d}$. It thus follows that

$$\mathbb{E}\left[g_{i,j,A}(\mathcal{X}_i, \mathcal{X}_i \cup \mathcal{P}_n)\right] \ge e^{-n\|f\|_\infty \theta_d(2(\ell_{i,j}+1)r_n)^d} \mathbb{E}[g_{i,j,A}(\mathcal{X}_i)]$$

$$\simeq e^{-n\|f\|_\infty \theta_d(2(\ell_{i,j}+1)r_n)^d} r_n^{d(i-1)} \mu_{i,j,A}.$$

Recall that for $i \ge k+2$ fixed, $j \le \binom{i}{k+1}$. It thus follows that since $nr_n^d \to 0$,

$$\mathbb{E}\left[\sum_{j>0} j X_{i,j,A}^P\right] \simeq \frac{n^i r_n^{d(i-1)}}{i!} \sum_{j=1}^{\binom{i}{k+1}} j\mu_{i,j,A},$$

and that in particular,

$$\mathbb{E}\tilde{\beta}_{k,A}^P \simeq \frac{n^{k+2} r_n^{d(k+1)}}{(k+2)!} \mu_{k+2,1,A}.$$

A similar approach is taken to compute the variance:

$$\mathbb{E}\left[(\tilde{\beta}_{k,A}^P)^2\right] = \mathbb{E}\left[\sum_{\mathcal{Y}\subseteq\mathcal{P}_n} \sum_{i,i'=k+2}^m \sum_{j,j'>0} jj' h_{i,j,A}(\mathcal{Y},\mathcal{P}_n) h_{i',j',A}(\mathcal{Y},\mathcal{P}_n)\right]$$

$$+ \mathbb{E}\left[\sum_{\ell=0}^m \sum_{\substack{\mathcal{Y},\mathcal{Y}'\subseteq\mathcal{P}_n \\ \mathcal{Y}\neq\mathcal{Y}'}} \sum_{i,i'=k+2}^m \sum_{j,j'>0} jj' h_{i,j,A}(\mathcal{Y},\mathcal{P}_n) h_{i',j',A}(\mathcal{Y}',\mathcal{P}_n) \mathbb{1}_{\{|\mathcal{Y}\cap\mathcal{Y}'|=\ell\}}\right].$$

For the first term, note that $h_{i,j,A}(\mathcal{Y},\mathcal{P}_n) h_{i',j',A}(\mathcal{Y},\mathcal{P}_n) = 0$ unless $i = i'$ and $j = j'$, because i is the number of vertices of $\mathcal{Y}$ and j is the kth Betti number of the complex it spans. This means the first term has in fact already been analyzed:

$$\mathbb{E}\left[\sum_{\mathcal{Y}\subseteq\mathcal{P}_n} \sum_{i,i'=k+2}^m \sum_{j,j'>0} jj h_{i,j,A}(\mathcal{Y},\mathcal{P}_n) h_{i',j',A}(\mathcal{Y},\mathcal{P}_n)\right]$$

$$= \mathbb{E}\left[\sum_{\mathcal{Y}\subseteq\mathcal{P}_n} \sum_{i=k+2}^m \sum_{j>0} j^2 h_{i,j,A}(\mathcal{Y},\mathcal{P}_n)\right] \simeq \frac{n^{k+2} r_n^{d(k+1)}}{(k+2)!} \mu_{k+2,1,A}.$$

For the second, observe first that the terms corresponding to $\ell \neq 0$ vanish: $h_{i,j,A}(\mathcal{Y},\mathcal{P}_n) h_{i',j',A}(\mathcal{Y}',\mathcal{P}_n) \equiv 0$ if $|\mathcal{Y}\cap\mathcal{Y}'| = \ell > 0$, because in that case neither $\mathcal{Y}$ nor $\mathcal{Y}'$ is a whole component. When $\ell = 0$, applying Theorem 2.4 yields

$$\mathbb{E}\left[\sum_{\mathcal{Y},\mathcal{Y}'\subseteq\mathcal{P}_n} h_{i,j,A}(\mathcal{Y},\mathcal{P}_n) h_{i',j',A}(\mathcal{Y}',\mathcal{P}_n) \mathbb{1}_{\{\mathcal{Y}\cap\mathcal{Y}'=\emptyset\}}\right]$$

$$= \frac{n^{i+i'}}{i!i'!} \mathbb{E}\left[jj' h_{i,j,A}(\mathcal{X}_i, \mathcal{X}_i\cup\mathcal{X}_{i'}\cup\mathcal{P}_n) h_{i',j',A}(\mathcal{X}_{i'}, \mathcal{X}_i\cup\mathcal{X}_{i'}\cup\mathcal{P}_n)\right],$$

where again $\mathcal{X}_i$ and $\mathcal{X}_{i'}$ are independent collections of i and i' i.i.d. points distributed according to f, respectively. Making use of (10) thus yields

$$\text{Var}\left[\tilde{\beta}_{k,A}^P\right]$$

$$= \mathbb{E}\left[\tilde{\beta}_{k,A}^P\right] + \sum_{i,i'=k+2}^m \sum_{j,j'\geq1} \frac{n^{i+i'} jj'}{i!i'!}\Bigg\{ \mathbb{E}\left[h_{i,j,A}(\mathcal{X}_i, \mathcal{X}_i\cup\mathcal{X}_{i'}\cup\mathcal{P}_n) h_{i',j',A}(\mathcal{X}_{i'}, \mathcal{X}_i\cup\mathcal{X}_{i'}\cup\mathcal{P}_n)\right]$$

$$- \mathbb{E}\left[h_{i,j,A}(\mathcal{X}_i, \mathcal{X}_i\cup\mathcal{P}_n)\right] \mathbb{E}\left[h_{i',j',A}(\mathcal{X}_{i'}, \mathcal{X}_{i'}\cup\mathcal{P}_n)\right]\Bigg\}.$$

9

Now, let $\mathcal{P}'_n$ be an independent copy of $\mathcal{P}_n$. Then

$$\mathbb{E}\left[h_{i,j,A}(\mathcal{X}_i, \mathcal{X}_i \cup \mathcal{X}_{i'} \cup \mathcal{P}_n)h_{i',j',A}(\mathcal{X}_{i'}, \mathcal{X}_i \cup \mathcal{X}_{i'} \cup \mathcal{P}_n)\right]$$
$$- \mathbb{E}\left[h_{i,j,A}(\mathcal{X}_i, \mathcal{X}_i \cup \mathcal{P}_n)\right]\mathbb{E}\left[h_{i',j',A}(\mathcal{X}_{i'}, \mathcal{X}_{i'} \cup \mathcal{P}_n)\right]$$

$$= \mathbb{E}\Bigg[h_{i,j,A}(\mathcal{X}_i, \mathcal{X}_i \cup \mathcal{X}_{i'} \cup \mathcal{P}_n)h_{i',j',A}(\mathcal{X}_{i'}, \mathcal{X}_i \cup \mathcal{X}_{i'} \cup \mathcal{P}_n)$$
$$- h_{i,j,A}(\mathcal{X}_i, \mathcal{X}_i \cup \mathcal{P}_n)h_{i',j',A}(\mathcal{X}_{i'}, \mathcal{X}_{i'} \cup \mathcal{P}'_n)\Bigg]$$

$$= \mathbb{E}\left[\left(h_{i,j,A}(\mathcal{X}_i, \mathcal{X}_i \cup \mathcal{X}_{i'} \cup \mathcal{P}_n) - h_{i,j,A}(\mathcal{X}_i, \mathcal{X}_i \cup \mathcal{P}_n)\right)h_{i',j',A}(\mathcal{X}_{i'}, \mathcal{X}_i \cup \mathcal{X}_{i'} \cup \mathcal{P}_n)\right]$$
$$+ \mathbb{E}\left[h_{i,j,A}(\mathcal{X}_i, \mathcal{X}_i \cup \mathcal{P}_n)\left(h_{i',j',A}(\mathcal{X}_{i'}, \mathcal{X}_i \cup \mathcal{X}_{i'} \cup \mathcal{P}_n) - h_{i',j',A}(\mathcal{X}_{i'}, \mathcal{X}_{i'} \cup \mathcal{P}_n)\right)\right]$$
$$+ \mathbb{E}\left[h_{i,j,A}(\mathcal{X}_i, \mathcal{X}_i \cup \mathcal{P}_n)\left(h_{i',j',A}(\mathcal{X}_{i'}, \mathcal{X}_{i'} \cup \mathcal{P}_n) - h_{i',j',A}(\mathcal{X}_{i'}, \mathcal{X}_{i'} \cup \mathcal{P}'_n)\right)\right]$$
$$=: E_1 + E_2 + E_3.$$

Now, observe that in fact $E_1 = 0$: the difference is non-zero if and only if $\mathcal{X}_i$ and $\mathcal{X}_{i'}$ are connected by an edge, in which case the second factor is zero.

Observe that the difference in E_2 is either 0 or -1. Furthermore, it is non-zero if and only if $\mathcal{X}_i$ and $\mathcal{X}_{i'}$ are connected by an edge, and both $\mathcal{X}_i$ and $\mathcal{X}_{i'}$ are connected. This probability is bounded above by

$$\|f\|_\infty^{i+i'-1}\theta_d^{i+i'-1}(2\ell_{i,j}r_n)^{d(i-1)}(2\ell_{i',j'}r_n)^{d(i'-1)}(2(\ell_{i,j} + \ell_{i'+j'} + 1)r_n)^d.$$

Finally, conditional on the event $\left[\cup_{x\in\mathcal{X}_i}B_{2\ell_{i,j}r_n}(x)\right] \cap \left[\cup_{x\in\mathcal{X}_{i'}}B_{2\ell_{i',j'}r_n}(x)\right] = \emptyset$, the two terms of E_3 have the same distribution by the spacial independence property of the Poisson process. A contribution from E_3 therefore only arises if in particular $\mathcal{X}_i$ and $\mathcal{X}_{i'}$ are both connected and the intersection above is non-empty. The probability of this event is bounded above by

$$\|f\|_\infty^{i+i'-1}\theta_d^{i+i'-1}(2\ell_{i,j}r_n)^{d(i-1)}(2\ell_{i',j'}r_n)^{d(i'-1)}(4(\ell_{i,j} + \ell_{i',j'})r_n)^d.$$

It follows that

$$\mathrm{Var}\left[\tilde{\beta}_{k,A}^P\right] = \mathbb{E}\left[\tilde{\beta}_{k,A}^P\right] + E,$$

and

$$|E| \leq \sum_{i,i'=k+2}^m \sum_{j,j'\geq 1} \frac{n^{i+i'}jj'(C_{i,j}r_n)^{d(i+i'-1)}}{i!i'!}2\|f\|_\infty^{i+i'-1}\theta_d^{i+i'-1} \leq C(f,d,k)(nr_n^d)^{k+2}(n^{k+2}r_n^{d(k+1)}),$$

where $C(f,k,d)$ is a constant depending on f, d, and k. This completes the proof of the first statement of the lemma. The proof of the second statement is the same, just removing the terms of the sum indexed by $i > k+2$, and the third statement is gotten by removing the terms indexed by $i = k+2$. $\square$

The following was proved via Stein's method in [2].

Theorem 2.7. *With notation as above, and for $n^{k+2}r_n^{d(k+1)} \to \infty$ and $nr_n^d \to 0$,*

$$\frac{S_k^P - \mathbb{E}\left[S_k^P\right]}{\sqrt{n^{k+2}r_n^{d(k+1)}}} \Longrightarrow \mathcal{N}\left(0, \frac{\mu_{k+2,1}}{(k+2)!}\right).$$

This gives a central limit theorem for $\tilde{\beta}_k^P$ as follows. Write

$$\tilde{\beta}_k^P = S_k^P + R_k^P, \qquad R_k^P := \sum_{\substack{i \geq k+3 \\ j \geq 1}} jX_{i,j}.$$

Fix $t \in \mathbb{R}$ and $\epsilon > 0$. By Lemma 2.6, for n large enough, $\frac{\mathrm{Var}(R_k^P)}{n^{k+2}r_n^{d(k+1)}} \leq \epsilon^3$, so that for n large enough,

$$\mathbb{P}\left[\frac{\tilde{\beta}_k^P - \mathbb{E}\tilde{\beta}_k^P}{\sqrt{n^{k+2}r_n^{d(k+1)}}} \leq t\right] \leq \mathbb{P}\left[\frac{S_k^P - \mathbb{E}S_k^P}{\sqrt{n^{k+2}r_n^{d(k+1)}}} \leq t + \epsilon\right] + \mathbb{P}\left[\left|\frac{R_k^P - \mathbb{E}R_k^P}{\sqrt{n^{k+2}r_n^{d(k+1)}}}\right| > \epsilon\right]$$

$$\leq \mathbb{P}\left[\frac{S_k^P - \mathbb{E}S_k^P}{\sqrt{n^{k+2}r_n^{d(k+1)}}} \leq t + \epsilon\right] + \epsilon.$$

By applying the central limit theorem already proved for S_k^P and then letting $\epsilon \to 0$, it follows that

$$\limsup_{n \to \infty} \mathbb{P}\left[\frac{\tilde{\beta}_k^P - \mathbb{E}\tilde{\beta}_k^P}{\sqrt{n^{k+2}r_n^{d(k+1)}}} \leq t\right] \leq \mathbb{P}\left[\sqrt{\frac{\mu_{k+2,1}}{(k+2)!}}Z \leq t\right].$$

The other inequality is proved in the same way, giving the following central limit theorem for $\tilde{\beta}_k^P$.

Theorem 2.8. *For notation as above,*

$$\frac{1}{\sqrt{n^{k+2}r_n^{d(k+1)}}}\left(\tilde{\beta}_k^P - \mathbb{E}\tilde{\beta}_k^P\right) \Longrightarrow \mathcal{N}\left(0, \frac{\mu_{k+2,1}}{(k+2)!}\right).$$

The remaining work is to use this result to obtain the same result for $\tilde{\beta}_k$ itself. To do so, the following "de-Poissonization result" is used.

Theorem 2.9 (Theorem 2.12 of [3]). *Suppose that for each $n \in \mathbb{N}$, $H_n(\mathcal{X})$ is a real-valued functional on finite sets $\mathcal{X} \subseteq \mathbb{R}^d$. Suppose that for some $\sigma^2 \geq 0$,*

(i) $\frac{1}{n}\mathrm{Var}(H_n(\mathcal{P}_n)) \longrightarrow \sigma^2$, *and*

(ii) $\frac{1}{\sqrt{n}}\left[H_n(\mathcal{P}_n) - \mathbb{E}H_n(\mathcal{P}_n)\right] \Longrightarrow \sigma^2 Z$, *for Z a standard normal random variable.*

Suppose that there are constants $\alpha \in \mathbb{R}$ and $\gamma > \frac{1}{2}$ such that the increments $R_{q,n} = H_n(\mathcal{X}_{q+1}) - H_n(\mathcal{X}_q)$ satisfy

$$\lim_{n \to \infty}\left(\sup_{n-n^\gamma \leq q \leq n+n^\gamma} |\mathbb{E}[R_{q,n}] - \alpha|\right) = 0, \tag{11}$$

$$\lim_{n \to \infty}\left(\sup_{n-n^\gamma \leq q < q' \leq n+n^\gamma} |\mathbb{E}[R_{q,n}R_{q',n}] - \alpha^2|\right) = 0, \tag{12}$$

and

$$\lim_{n \to \infty}\left(\frac{1}{\sqrt{n}}\sup_{n-n^\gamma \leq q \leq n+n^\gamma} \mathbb{E}[R_{q,n}^2]\right) = 0. \tag{13}$$

Finally, assume that there is a constant $\tilde{\beta} > 0$ such that, with probability one,

$$|H_n(\mathcal{X}_q)| \leq \tilde{\beta}(n+q)^\beta.$$

11

Then $\alpha^2 \leq \sigma^2$ and as $n \to \infty$, $\frac{1}{n}\mathrm{Var}(H_n(\mathfrak{X}_n)) \to \sigma^2 - \alpha^2$ and

$$\frac{1}{\sqrt{n}}\big[H_n(\mathfrak{X}_n) - \mathbb{E}H_n(\mathfrak{X}_n)\big] \implies \sqrt{\sigma^2 - \alpha^2}Z.$$

In conjunction with Theorem 2.7, this yields Theorem 2.1, as follows.

Proof of Theorem 2.1. Theorem 2.9 is applied to the functional

$$H_n(\mathfrak{X}) := \frac{1}{\sqrt{(nr_n^d)^{k+1}}} \sum_{\mathcal{Y} \subseteq \mathfrak{X}} \left(\sum_{i=k+2}^{m} \sum_{j \geq 1} j h_{i,j}(\mathcal{Y}, \mathfrak{X}) \right);$$

$\sigma^2 = \frac{\mu_{k+2,1}}{(k+2)!}$ and the central limit theorem holds for $H_n(\mathcal{P}_n)$ by Theorem 2.8.
Let

$$D_{q,n} := \sum_{\mathcal{Y} \subseteq \mathfrak{X}_{q+1}} \left(\sum_{i=k+2}^{m} \sum_{j \geq 1} j h_{i,j}(\mathcal{Y}, \mathfrak{X}_{q+1}) \right) - \sum_{\mathcal{Y} \subseteq \mathfrak{X}_q} \left(\sum_{i=k+2}^{m} \sum_{j \geq 1} j h_{i,j}(\mathcal{Y}, \mathfrak{X}_q) \right)$$

(the dependence on n is only through the threshhold radius r_n), and observe that $D_{q,n}$ is the kth Betti number of the component of X_{q+1} in $\mathfrak{X}_{q+1}$, minus the kth Betti number of the complex that results by taking the component of X_{q+1} and removing X_{q+1} from it, assuming these components are on m or fewer vertices. It follows that the difference is bounded by $\binom{m}{k+1}$, and is only non-zero if X_{q+1} is connected to at least $k+1$ other vertices, so that

$$|\mathbb{E}[D_{q,n}]| \leq \binom{m}{k+1}\binom{n+n^\gamma}{k+1} \left(\|f\|_\infty r_n^d\right)^{(k+1)} \leq \binom{m}{k+1} \left((n+n^\gamma)\|f\|_\infty r_n^d\right)^{(k+1)}.$$

The first condition of the theorem is then satisfied with $\alpha = 0$, for any $\gamma \in \left(\frac{1}{2}, 1\right]$.

Next, consider the quantity $\mathbb{E}[D_{q,n}D_{q',n}]$ for $q < q'$. By the observation above, $D_{q,n}D_{q',n}$ is uniformly bounded by $\binom{m}{k+1}^2$. The probability that $D_{q,n} \neq 0$ is bounded above by $c[n+n^\gamma]^{k+1}r_n^{d(k+1)}$ as before. Given that that difference is non-zero, the largest probability event that causes $D_{q',n}$ to be non-zero is that $X_{q'+1}$ is connected to the component of X_{q+1}, and that its removal changes the Betti number of that component. The probability that $X_{q'+1}$ is in the component of X_{q+1} is bounded above cr_n^d, so that

$$|\mathbb{E}[R_{q,n}R_{q',n}]| = \frac{1}{(nr_n^d)^{k+1}}|\mathbb{E}[D_{q,n}D_{q',n}]| \leq \frac{1}{(nr_n^d)^{k+1}}\left(c[n+n^\gamma]^{k+1}r_n^{d(k+2)}\right) \leq cr_n^d,$$

so that the second condition of the theorem is also satisfied.

If $q = q'$, then we have as above

$$\mathbb{E}[D_{q,n}^2] \leq c[n+n^\gamma]^{k+1}r_n^{d(k+1)},$$

so that

$$\frac{1}{\sqrt{n}}\mathbb{E}[R_{q,n}^2] \leq \frac{2^{k+1}c}{\sqrt{n}},$$

and so the third condition is satisfied as well.

Finally, the polynomial boundedness condition of Theorem 2.9 is satisfied trivially, and the proof is complete.

$\square$

References

[1] Matthew Kahle. Sharp vanishing thresholds for cohomology of random flag complexes. to appear in Ann. of Math., 2014.

[2] Matthew Kahle and Elizabeth Meckes. Limit theorems for Betti numbers of random simplicial complexes. *Homology Homotopy Appl.*, 15(1):343–374, 2013.

[3] Mathew Penrose. *Random geometric graphs*, volume 5 of *Oxford Studies in Probability*. Oxford University Press, Oxford, 2003.